ROMAI J., **3**, 2(2007), 21–30

NUMERICAL ALGORITHMS FOR EVALUATING THE TRAFFIC COEFFICIENT IN GENERALIZED QUEUEING SYSTEMS

Olga Benderschi, Gheorghe Mishkoy, Nicolae Andronati, Ioan Griza

State University of Moldova, Academy of Sciences of Moldova, Free International University of Moldova, Chişinău, Republica Moldova

obenderschi@usm.md

Abstract The traffic coefficient is an important measure of a queueing system's performance and it is responsible for the workload of the system. Analysis of queueing systems delivers formulae for system performance characteristics - many of such analytical expressions involve the traffic coefficient ρ. In the case of queueing systems $M|G|1|\infty$ and $M_r|G_r|1|\infty$ with priorities one can easily evaluate ρ via analytic formulae using rates of incoming flows and mean values of corresponding service times. However, in the case of priority systems with random switchover times, one should be able to evaluate the Laplace-Stieltjes transforms (LST's) of system busy periods in order to estimate the value of the traffic coefficient. Generally, this can only be done numerically.

Algorithms of evaluation of the traffic coefficient in general queueing systems are presented. Values of the traffic coefficient for some given values of the systems' parameters are also calculated.

1. INTRODUCTION

It is a usual practice to represent and study real world phenomena processes by mathematical models. Among such models there are models of priority queueing systems. The theory of priority queueing systems is concerned with the phenomena of prioritised servicing - the incoming requests should be classified by their importance and served according to the set priority levels. In comparison with other queueing models the priority queueing systems have

a more complicated structure, which limits the possibility of their exact analytical analysis. Thus, many results are derived considering the stationary behaviour of the system. This method has no alternatives in the case when the exact analysis is not possible.

2. PRIORITY QUEUEING SYSTEMS WITH SWITCHOVER TIMES

2.1. DESCRIPTION

Consider a queueing system with a single station and r classes of incoming requests, each having its own flow of arrival and waiting line. We call the requests from i^{th} queueing line L_i i-requests. i-Requests have a higher priority than j-requests if $1 \leq i < j \leq r$. The station gives a preference in service to the requests of the highest priority among those presented in the system.

Suppose that the time periods between two consecutive arrivals of the requests of the class i are independent and identically distributed with some common cumulative distribution function (cdf) $A_i(t)$ with mean $\mathbb{E}[A_i]$, $i = 1, \ldots, r$. Similarly, suppose that service time of a customer of the class i is a random variable B_i with a cumulative distribution function $B_i(t)$ with mean service time $\mathbb{E}[B_i]$, $i = 1, \ldots, r$.

However, some time is needed for server to proceed with the switching from one line of requests to another. This time is considered to be a random variable, and we say that C_{ij} is the time of switching from the service of i-requests to the service of j-requests, if $1 \leq i, j \leq r$, $i \neq j$.

A large class of priority queueing systems can potentially be described using the following information and identifiers:

- arrival flows - distributions of inter-arrival times (for each flow);
- service times - distributions of service times (for each flow);
- switching times - specification of the switching type (neutral state or not) and distributions of switching times;

- warming time - distribution of warming times;
- order of service within a line (FIFS, LIFS, RANDOM);
- service discipline;
- switching discipline;
- behaviour of the server in the idle state.

We adopt classification and terminology introduced in [1,4]. We also should explain some additional notions and notations.

Definition 1. *By kk-busy period we call the period of time which starts when a k-request enters the empty system and finishes when there are no longer k-requests in the system. Denote the kk-busy period by Π_{kk}.*

Definition 2. *By k-busy period we call the period of time which starts when an i-request enters the empty system, $i \leq k$, and finishes when there are no longer k-requests in the system. Denote the k-busy period by Π_k.*

Note, that r-busy period is nothing else but the system's busy period Π, i.e. $\Pi \equiv \Pi_r$.

The following two notions are due to the fact that both servicing and switching can be interrupted under preemptive service and switching policies.

Definition 3. *By k-cycle of service we call the period of time which starts when server begins the servicing of a k-request, and finishes when this request leaves the system. Denote the k-cycle of service by H_k. If the servicing of a certain k-request is not interrupted then the corresponding realisation of H_K coincides with the time B_k this request was being serviced.*

Definition 4. *By k-cycle of switching we call the period of time which starts when server begins the switching to the line of k-requests, and finishes when server is ready to provide service to these requests. Denote the k-cycle of switching by H_k. If the servicing of a certain k-request is not interrupted, then*

the corresponding realisation of N_k coincides with the time B_k this request was being serviced. If the switching from i^{th} line to the k^{th} line is not interrupted, then the corresponding realisation of N_k coincides with the time C_{ik} this ik-switching lasted.

Let $\Pi_{kk}(t)$, $\Pi_k(t)$, $H_k(t)$ and $N_k(t)$ be the cumulative distribution functions of kk-busy periods, k-busy periods, k-cycle of service and k-cycle of switching, respectively. Let also $\pi_{kk}(t)$, $\pi_k(t)$, $h_k(t)$ and $\nu_k(t)$ be their Laplace-Stieltjes transform, i.e.

$$\pi_{kk}(s) = \int_0^\infty e^{-st} d\Pi_{kk}(t), \ldots, \ \nu_k(s) = \int_0^\infty e^{-st} dN_k(t).$$

Finally, let $\beta_k(s)$ be the Laplace-Stieltjes transform of $B_i(t)$, i.e.

$$\beta_i(s) = \int_0^\infty e^{-st} dB_i(t).$$

From now on we assume that C_{ij} do not depend on i and depend only on j, i.e. $C_{ij} \equiv C_j \ \forall j = 1, \ldots, r$. Denote the Laplace-Stieltjes transform of C_j with cdf $C_j(t)$ by $c_j(s)$:

$$c_j(s) = \int_0^\infty e^{-st} dC_j(t).$$

2.2. PRIORITY QUEUEING SYSTEMS WITH POISSON INCOMING FLOWS

The queueing systems with Poisson incoming flows are of great importance in the theory and practice. In this case the inter-arrival times are exponentially distributed, i.e. $A_i(t) = 1 - e^{-a_i t}$, $i = 1, \ldots, r$, where a_1, a_2,..., a_r are some positive real numbers with the physical meaning of the flow arrival rates. The compound flow of the flows of 1-, 2-, ...,k-requests is Poisson with arrival rate $\sigma_k = \sum_{i=1}^{k} a_i$.

Adopting and slightly extending the standard Kendall notation we write $M_r|G_r|1|\infty$ to denote a priority queueing system with Poisson incoming flows of requests and random switchover times.

3. TRAFFIC COEFFCIENT AND ITS CALCULATION

Analysis of queueinq systems delivers formulae for systems performance characteristic—many of such analytical expressions involve the traffic coefficient ρ. In the case of priority queueing systems $M_r|G_r|1|\infty$ with degenerated times of orientation one can easily evaluate ρ via analytic formulae using rates of incoming flows and mean values of corresponding service times.

For instance, the traffic coefficient of an $M_r|G_r|1|\infty$ system can be calculated as follows [3]

$$\rho = \sum_{i=1}^{r} a_i b_i, \tag{1}$$

where

- for the scheme "non-identical servicing again" $b_i = \frac{1}{\sigma_{i-1}} \left[\frac{1}{\beta_i(\sigma_{i-1})} - 1 \right]$;
- for the scheme "resume" $b_i = \mathbb{E}[B_i]$;
- for the scheme "with losses" $b_i = \frac{1}{\sigma_{i-1}} \left[1 - \beta_i(\sigma_{i-1}) \right]$.

In the case when $\rho > 1$, the following takes place: $\pi(0) < 1$, and $\Pi(t)$ is an improper cumulative distribution function, i.e.

$$\lim_{t \to \infty} \Pi(t) < 1,$$

which means that a busy period is of indefinite length with a positive probability. However, if $\rho < 1$, then $\pi(0) = 1$ and the cdf $\Pi(t)$ of the busy period Π is proper. These comments motivate the presence of $\pi(0) \equiv \pi_r(0)$ in the following examples.

Example 1. *Consider the system $M_{10}|M_{10}|1|\infty$. In this case*

$$\beta_i(s) = \frac{1}{s\mathbb{E}[B_i] + 1},\ i = 1, \ldots, 10.$$

Let $a_i = 1$ *and* $\mathbb{E}[B_i] = 0.05$, $i = 1, \ldots, 10$. *Let also* $\epsilon = 0.001$. *The results of our calculations of* ρ *can be found in Table 1.*

schemes:	*non-identical servicing again*	*resume*	*with losses*
ρ	0.5	0.5	0.413914
$\pi_{10}(0)$	0.999959	0.999959	0.999989

Table 1 Calculation of the traffic coefficient ρ and $\pi_{10}(0)$ in Example 1.

Example 2. *Consider* $M_{10}|G_{10}|1|\infty$ *under the scheme "non-identical servicing again" where the service times of the requests from the queueing priority lines* L_1, L_2, L_3, L_4 *are exponential* $Exp(20)$, *the service times of the requests from the lines* L_5, L_6, L_7, L_8 *are uniformly distributed on [0,1], and, finally, the service times of the requests from the lines* L_9 *and* L_{10} *are of Erlang type* $Er(2, 20)$. *In this case we have*

$$\beta_i(s) = \frac{20}{20+s},\ i = 1, 2, 3, 4,$$
$$\beta_i(s) = 1 - e^{-s},\ i = 5, 6, 7, 8,$$
$$\beta_i(s) = \left(\frac{20}{20+s}\right)^2,\ i = 9, 10.$$

For this system the traffic coefficient ρ *was numerically estimated to be equal to* $104.44 \gg 1$, *whereas* $\pi(0) \equiv \pi_{10}(0) = 0.48 < 1$. *This clearly shows that the system is under the heavy traffic regime.*

In what follows we present algorithms of numerical evaluation of the LST of the k-busy periods and the traffic coefficient ρ. For simplicity we give the algorithms for the systems $M_r|G_r|1|\infty$ under the scheme "non-identical servicing again" and preemptive switching policy. One needs to be able to

evaluate the LST of the busy period (r-busy period) in order to calculate ρ. This can be done using the following

Algorithm 1 ($M_r|G_r|1$ *systems with random switchings under pre-emptive non-identical servicing again*)

Input: r, s^*, $\epsilon > 0$, $\{a_i\}_{i=1}^r$, $\{\beta_i(s)\}_{i=1}^r$, $\{c_i(s)\}_{i=1}^r$.

Output: $\pi_k(s^*)$

Description:

IF (k==0) THEN $\pi_0(s^*) := 0$; *RETURN*

$k := 1$; $q := 1$; $\sigma_0 := 0$;

Repeat

inc(q);

$\sigma_q := \sigma_{q-1} + a_q$;

Until $q == r$;

Repeat

$\nu_k(s) := c_k(s^* + \sigma_{k-1})\{1 - \frac{\sigma_{k-1}}{s^*+\sigma_{k-1}}\left[1 - c_k(s^* + \sigma_{k-1})\right]\pi_{k-1}(s^*)\}^{-1}$;

$h_k(s^*) := \beta_k(s + \sigma_{k-1})\{1 - \frac{\sigma_{k-1}}{s^*+\sigma_{k-1}}\left[1 - \beta_k(s^* + \sigma_{k-1})\right]\pi_{k-1}(s^*)\nu_k(s^*)\}^{-1}$;

$\pi_{kk}^{(0)}(s^*) := 0$; $n := 1$;

Repeat

$\pi_{kk}^{(n)}(s^*) := h_k(s^* + a_k - a_k\pi_{kk}^{(n-1)})$;

inc(n);

Until $|\pi_{kk}^{(n)}(s^*) - \pi_{kk}^{(n-1)}(s^*)| < \epsilon$;

$\pi_k(s^*) := \frac{\sigma_{k-1}\pi_{k-1}(s^*+a_k)}{\sigma_k} + \frac{\sigma_{k-1}}{\sigma_k}(\pi_{k-1}(s^* + a_k - a_k\pi_{kk}(s^*))) - \pi_{k-1}(s^* + a_k)\nu_k(s^* + a_k[1 - \pi_{kk}(s^*)]) + \frac{a_k}{\sigma_k}\nu(s^* + a_k - a_k\pi_{kk}(s^*))\pi_{kk}(s^*)$;

inc(k);

Until $k == r$;

End of Algorithm 1.

Remark 1. *The algorithm 1 is convergent but it does not provide one with an absolute error of the approximation. In this algorithm some small quantity* ϵ *is used to judge on the convergence of the Cauchy sequence* $\{\pi_{kk}^{(n)}(s^*)\}_{n=0}^{\infty}$.

Next we present the algorithm of calculation of the traffic coefficient.

Algorithm 2 ($M_r|G_r|1$ ***systems with random switchings under pre-emptive non-identical servicing again*****)**

Input: r, $\{a_i\}_{i=1}^r$, $\{\beta_i(s)\}_{i=1}^r$, $\{c_i(s)\}_{i=1}^r$.

Output: ρ

Description:

$k := 1$; $\rho := 1$; $\sigma_0 := 0$; $\sigma_1 := a_1$;

$f_1 := 1$; $p := 1$;

$b_1 := -(\beta'(0) + c_1'(0))/(1 - a_1 c_1'(0))$;

$\rho := a_1 b_1$;

Repeat

$inc(k)$;

$\sigma_k := \sigma_{k-1} + a_k$;

$b_k := p\frac{1}{\sigma_{k-1}c_k(\sigma_{k-1})}\left(\frac{1}{\beta_k(\sigma_{k-1})} - 1\right)$;

$\rho := \rho + a_k b_k$;

$f_k := 1 + \frac{\sigma_k - \sigma_{k-1}\pi_{k-1}(a_k)}{\sigma_{k-1}}\left(\frac{1}{c_k(\sigma_{k-1})} - 1\right)$;

$p := f_k p$;

Until $k == r$;

End of Algorithm 2.

Remark 2. *Calculation of* ρ *in Algorithm 2 requires calculation of* $\pi_{k-1}(a_i)$, $k = 2, \ldots, r$, *which, in turn, can be realised using the algorithm Algorithm 1. For a different scheme of servicing or switching policy one should employ the corresponding formulae for the LST's* $h_k(s)$, $\nu_k(s)$, $\pi_{kk}(s)$, $\pi_k(s)$ *(see [3]).*

Example 3. *Consider the system* $M_8|M_8|1$ *with interarrival times* $Exp(1)$ *and exponential service times* $Exp(40)$. *The switchover times* C_k *are all distributed as* $Exp(200)$, $k = 1, \ldots, 8$. *The results of calculations for such systems can be found in Table 2. The quantity* ϵ *was taken to be* 0.001, *yet one can notice that, although* $\pi(0) \equiv \pi_8(0)$ *is very close to 1 in all three cases (as it should be, because* $\rho < 1$*), the absolute precision for the first two schemes is less than* ϵ, *see Remark 1.*

schemes:	*preemptive servicing again with losses*	*non-identical servicing again*	*preemptive resume servicing*
ρ	0.194987	0.211303	0.211308
$\pi_8(0)$	0.982491	0.980877	0.999970

Table 2 Calculation of the traffic coefficient ρ and $\pi_8(0)$ in Example 2.

All schemes employ preemptive switching policy.

4. CONCLUSIONS

We presented a model algorithm of the numerical evaluation of the traffic coefficient in generalised priority queueing systems (Algorithm 2). This algorithm makes use of the LST of busy period of the system—this should also be calculated numerically, for instance, using Algorithm 1. However, we found that (i) the number of priority flows r should not exceed 10-12 for satisfactory fast calculations, and, (ii) the calculation of the LST of busy periods is performed without clear idea about the absolute error of the evaluation. Therefore, there is a necessity of further optimisation of these numerical methods in order to achieve high level of precision and allow to consider greater number of priority waiting lines. Such work is being done currently [2,5].

5. ACKNOWLEDGEMENT

This work was done under support of the SCOPES grant IB7320-110720 and RFFI grant 0644CRF.

Authors also express their gratitude to Andrei Bejan for the help in preparation of this paper.

References

[1] Gh. Mishkoy, S. Giordano, A. Bejan, O. Benderschi, *Priority queueing systems with switchover times: Generalized models for QoS and CoS network technologies*, Comput. Sci. J. Moldova, **15**(**44**), 2, 217-242.

[2] G. Mishkoĭ, V. Rykov, S. Giordano, A. Bezhan, *Mnogomernye analogi uravneniya Kendalla i zagruzka pribora dlya prioritetnyh sistem: vychislitel'nye aspekt*, Avtomatica i Telemekhanika, 2006, accepted for publication, in Russian.

[3] G. P. Klimov, G. K. Mishkoĭ, *Prioritetnye sistemy obsluzhivaniya s orientatsieĭ*, M.: Izdatel'stgo Moscovskovo Universiteta, 1979. (Russian)

[4] G. Mishkoy, S. Giordano, N. Andronati, A. Bejan, *Priority queueing systems with switchover times: generalized models for QoS and CoS network technologies and analysis*, Technical report, 2006. WEB: http://www.vitrum.md/andrew/PQSST.pdf

[5] A. Bejan, *Numerical treatment of the Kendall equation in the analysis of priority queueing systems*, Bul. Acad. Şt. Rep. Mold. Mat., **51**(2) (2006), 17-28.